15 The Hamilton–Jacobi–Bellman equation

We begin a study of deterministic continuous-time controllable dynamical systems with a heuristic derivation of the Hamilton–Jacobi–Bellman equation. Then we prove that any suitably well-behaved solution of this equation must coincide with the infimal cost function and that the minimizing action gives an optimal control.

Recall from Subsection 1.3 that a continuous-time controllable dynamical system is a map

$$b : \mathbb{R}^+ \times \mathbb{R}^d \times A \to \mathbb{R}^d.$$

We now assume that the action-space A is a subset of $\mathbb{R}^p$ for some p, in examples A is often simply an interval in $\mathbb{R}$. We assume also that b is continuous, and is differentiable in x with bounded derivative. A control is a map $u : \mathbb{R}^+ \to A$. Given a control u and a starting time and state (s, x), we define[26] the controlled path $(x_t)_{t\geqslant s}$ as the solution of the differential equation

$$\dot{x}_t = b(t, x_t, u_t), \quad t \geqslant s, \quad x_s = x.$$

We shall consider two types of optimization problem. In the first type, we fix a *stopping set* $D \subseteq \mathbb{R}^d$ and a *time-horizon* $T < \infty$ and specify continuous and bounded *cost functions*[27]

$$c : [0, T) \times \mathbb{R}^d \times A \to \mathbb{R}, \quad C : \{T\} \times D \to \mathbb{R}.$$

We say that a control u is *feasible*, starting from (s, x), if, for the associated controlled path starting from (s, x), we have $x_T \in D$. If there is no such control, then we say (s, x) is *infeasible*. In the second type of problem, we also fix a stopping set $D \subseteq \mathbb{R}^d$, which is the boundary of some open set $S \subseteq \mathbb{R}^d$, but the time of arrival in D is *unconstrained*. We specify continuous and bounded cost functions

$$c : \mathbb{R}^+ \times S \times A \to \mathbb{R}, \quad C : \mathbb{R}^+ \times D \to \mathbb{R}.$$

We say that a control u is *feasible*, starting from (s, x), if $\tau < \infty$, where

$$\tau = \inf\{t \geqslant 0 : x_t \in D\}.$$

In order to give a unified treatment of the two cases, we shall, in the first case, set $\tau = T$ and write $\tilde{S} = ([0, T) \times \mathbb{R}^d)$ and $\tilde{D} = \{T\} \times D$. In the the second case, we write $\tilde{S} = \mathbb{R}^+ \times S$ and $\tilde{D} = \mathbb{R}^+ \times D$.

The *total cost* for a feasible control u, starting from $(s, x) \in \tilde{S}$, is defined by

$$V^u(s, x) = \int_s^\tau c(t, x_t, u_t)dt + C(\tau, x_\tau).$$

The *infimal cost function* V is defined by

$$V(s, x) = \inf_u V^u(s, x),$$

[26] The basic theory of existence and uniqueness for solutions of differential equations is reviewed, and its application in this setting is explained, in Section 18.

[27] As usual, any problem of maximizing rewards can be treated as a problem of minimizing negative costs, so we do not discuss the theory for this sort of problem separately.

where the infimum is taken over all continuous feasible controls starting from (s, x), and $V(s, x) = \infty$ if there are no such controls.

Suppose we start from $(t, x) \in \tilde{S}$ and choose action a until a short time later $t + \delta$, then switching to an optimal control. Comparing this control with the optimal control from (t, x), we obtain, up to terms which are small compared to δ,

$$V(t, x) \leqslant c(t, x, a)\delta + V(t + \delta, x + b(t, x, a)\delta)$$

On the other hand, by optimizing the right-hand side over a we might expect to get arbitrarily close to $V(t, x)$. We expand to first order

$$V(t + \delta, x + b(t, x, a)\delta) = V(t, x) + \dot{V}(t, x)\delta + \nabla V(t, x)b(t, x, a)\delta + O(\delta^2).$$

On substituting this in the inequality, rearranging, dividing by δ and letting $\delta \to 0$, we obtain

$$\inf_a \{c(t, x, a) + \dot{V}(t, x) + \nabla V(t, x)b(t, x, a)\} = 0, \quad (t, x) \in \tilde{S}.$$

This is called the *Hamilton–Jacobi–Bellman equation.* It is the optimality equation for continuous-time systems. The final cost C provides a boundary condition $V = C$ on $\tilde{D}$.

Proposition 15.1. *Suppose that there exists a function $F : \tilde{S} \cup \tilde{D} \to \mathbb{R}$, differentiable with continuous derivative, and that, for a given starting point $(s, x) \in \tilde{S}$, there exists a continuous feasible control u^* such that*

$$c(t, x, a) + \dot{F}(t, x) + \nabla F(t, x)b(t, x, a) \geqslant 0$$

for all $(t, x) \in \tilde{S}$ and $a \in A$, with equality when $t \in [s, \tau^*)$ and $(x, a) = (x^*_t, u^*_t)$. Suppose also that $F = C$ on $\tilde{D}$. Then $F(s, x) = V(s, x)$ and u^* defines an optimal control starting from (s, x).

Proof. It will suffice to consider the case $s = 0$. Fix any continuous feasible control $u : \mathbb{R}^+ \to A$ and set

$$m_t = \int_0^t c(s, x_s, u_s)ds + F(t, x_t), \quad 0 \leqslant t \leqslant \tau.$$

Then m is continuous on $[0, \tau]$ and differentiable on $[0, \tau)$, with

$$\dot{m}_t = c(t, x_t, u_t) + \dot{F}(t, x_t) + \nabla F(t, x_t)b(t, x_t, u_t) \geqslant 0,$$

and with equality if $u = u^*$. Therefore

$$F(0, x) = m_0 \leqslant m_\tau = \int_0^\tau c(s, x_s, u_s)ds + C(\tau, x_\tau) = V^u(0, x),$$

with equality if $u = u^*$. □

The proposition sets up a possible way to calculate the infimal cost function and to find an optimal control. One tries to solve the Hamilton–Jacobi–Bellman equation

$$\inf_a \{c(t, x, a) + \dot{V}(t, x) + \nabla V(t, x)b(t, x, a)\} = 0, \quad (t, x) \in \tilde{S},$$

and to identify, for each $(t,x) \in \tilde{S}$ a minimizing action $u(t,x)$. Then, given a starting point $(s,x) \in \tilde{S}$, we attempt to solve the differential equation $\dot{x}_t = b^u(t,x_t)$, where $b^u(t,x) = b(t,x,u(t,x))$ and check that $\tau < \infty$ and $x_\tau \in D$. The control $u^*_t = u(t,x_t)$ then has $(x_t)_{s \leqslant t \leqslant \tau}$ as its controlled process starting from (s,x), so u^* has the minimizing property required by the proposition. In this case, we say that the function u *defines a feasible control for starting point* (s,x). It is often the case that the minimizing function $u(t,x)$ depends *discontinuously* but *piecewise continuously* on (t,x), and so do the associated controls. It is not hard to extend the proposition to this case, though we will not give details. In practice, the main hope to solve the HJB equation is to guess its shape as a function of x, to find the minimizing action $u(t,x)$ explicitly, and thereby to reduce the problem to a differential equation in t. These steps are illustrated in the next two examples.

Example (Linear system with quadratic costs). Consider the linear system with state-space $\mathbb{R}^d$ and action-space $\mathbb{R}^p$ given by $b(x,a) = Ax + Ba$, where A and B are matrices of appropriate dimensions. Take as cost function the non-negative quadratic function $c(x,a) = x^T Rx + a^T Qa$, which we shall assume to vanish only if $x = 0$ and $a = 0$. Suppose the final cost is also quadratic and non-negative, thus $C(x) = x^T \Pi(T) x$, for some matrix $\Pi(T)$.

As in the discrete-time case, let us try in the HJB equation a solution of the form $V(t,x) = x^T \Pi(t) x$, for some non-negative definite matrices $\Pi(t)$. We have

$$\begin{aligned} &\inf_a \{c(x,a) + \dot{V}(t,x) + \nabla V(t,x) b(x,a)\} \\ &\quad = \inf_a \{x^T(R + \Pi A + A^T \Pi + \dot{\Pi})x + x^T \Pi B a + a^T B^T \Pi x + a^T Q a\} = x^T(\tilde{R} - \tilde{S}^T Q^{-1} \tilde{S})x \end{aligned}$$

at $a = -Q^{-1}\tilde{S}x$, where $\tilde{R} = R + \Pi A + A^T \Pi + \dot{\Pi}$ and $\tilde{S} = B^t \Pi$. (See Section 10.) Hence V is a solution if and only if $(\Pi(t))_{0 \leqslant t \leqslant T}$ satisfies the continuous-time *Riccati equation*

$$\dot{\Pi} + R + \Pi A + A^T \Pi - \Pi B Q^{-1} B^T \Pi = 0.$$

Example (Managing investment income). The following may be considered as a model for optimizing utility from investment income over a prescribed lifetime T. We seek to maximize

$$\int_0^T e^{-\alpha s} \sqrt{u_s} ds$$

subject to $\dot{x}_t = \beta x_t - u_t$ and $x_t \geqslant 0$ for all $0 \leqslant t \leqslant T$. Thus α is the personal discount rate, β is the rate of interest, and $\sqrt{u}$ is the utility gained from income at rate u.

The optimality equation is

$$\sup_a \{e^{-\alpha t}\sqrt{a} + \dot{V}(t,x) + (\beta x - a)V'(t,x)\} = 0, \quad 0 \leqslant t \leqslant T,$$

with boundary condition $V(T,x) = 0$. By scaling, the maximal reward function must have the form

$$V(t,x) = e^{-\alpha t}\sqrt{v(t)x}$$

for some function v. By substitution in the optimality equation we obtain $\dot{v} - (2\alpha - \beta)v + 1 = 0$ with maximizing action $a = x/v(t)$. Hence

$$v(t) = \frac{1 - e^{-(2\alpha-\beta)(T-t)}}{2\alpha - \beta}$$

and the optimal control is $u_t = x_t/v(t)$.

A short-cut is available for this example using the Cauchy-Schwarz inequality. We have the constraint

$$0 = e^{-\beta T} x_T = x_0 - \int_0^T e^{-\beta s} u_s ds.$$

By Cauchy-Schwarz,

$$\int_0^T e^{-\alpha s} \sqrt{u_s} ds = \int_0^T (e^{-\beta s} u_s)^{1/2} (e^{-(2\alpha - \beta)s})^{1/2} ds \leqslant \sqrt{x_0} \left(\int_0^T e^{-(2\alpha-\beta)s} ds \right)^{1/2},$$

which confirms our calculation of $V(0, x_0)$.